

Mehring-Tóth Szilvia

## Replikázás a „Big Data és szociológia” kapcsán, avagy a módszertani paradigma és egy PhD-kutatás véletlen találkozása

### Bevezetés

A *Replika* 92–93. száma tematikus megközelítéssel több tanulmányt közölt, melyek a Big Data jelenségét, módszertani határait, tudományos alkalmazhatóságát elemzik. A publikációk sorozatában, mint egy jól összeállított SWOT-analízisben, a módszertani paradigma lehetőségeit és gyengeségeit ismerhetjük meg. A pró és kontra megközelítések hangvétele kifejezetten érdekes és olvasmányos, annak ellenére, hogy egy komoly módszertani kérdéskört mutatnak be. Bevallom, akkor a kötet számomra az első találkozás élményét jelentette a Big Data koncepciójával, és nagy lendülettel húztam alá azon részeket, melyek a kétségek oldaláról közelítették meg a kérdéskört. Ezzel egy időben a PhD-kutatásom adatgyűjtésén és módszertanának kialakításán dolgoztam a települési önkormányzatok képviselőinek témakörében, amikor sikerült hozzájutnom a [valasztas.hu](http://valasztas.hu) adatbázisának teljes digitalizált „tartalmához”. E véletlen találkozás nem várt közelségbe hozta a Big Data elméletét, és alaposan felülírta az első olvasás során kialakult álláspontomat. A következő sorokban a téma meghatározó szakértőinek elméleti álláspontját foglalom össze röviden, majd szembesítem azokat a saját gyakorlati tapasztalataimmal. Mivel álláspontom egy speciális adatbázis használata során formálódott, általános érvényű igazságokat meg sem próbálok megfogalmazni, csupán a Big Datával kapcsolatban a munka közben tett tapasztalataimat tudom leírni.

Az idézett *Replika*-szám hét tanulmánya tematikus egységbe szerkesztve, társadalomtudományi szempontból megközelítve gyűjtött össze kutatói véleményeket a napjaink digitális világában „keletkező” adatbázisok tudományos célú felhasználásáról. Az írásokat csoportosítva feltűnik, hogy Dessewffy és Láng vitaindító „tudományos kiáltványa”, valamint Csepeli György tanulmánya a „nagy adatokban” rejlő pozitívumokat, lehetőségeket hangsúlyozza. Mindkét írás szélesebb értelemben fogja fel a Big Datát, amikor a „digitális adatok robbanásszerű mennyiségi növekedése miatt előálló minőségileg is új társadalmi gyakorlatok és megismerési módok együttese”-ként határozza meg (Dessewffy és Láng 2015: 157). Dessewffy és Láng szerint az új ismeretelméleti paradigma nemcsak a 21. században keletkező megnövekedett adatmennyiség következtében alakult ki, a keletkezésnek meghatározó eleme volt a digitális rögzítések általános elterjedése is. E folyamatok által a megfigyelés eszköze változott meg, és eddig a társadalmi valóságból nem ismert mintázatok, rejtett működési elvek megismerése vált lehetővé. Ezek a „keletkező” adathalmazok ugyanis az emberi viselkedés nyomait rögzítik és követik, a „természeti és társadalmi folyamatok korábban elképzelhetetlen kiterjedésű és komplexitású modellezését teszi[k] lehetővé” (Csepeli 2015: 171). Az ilyen Big Data típusú adatbázisok előnyét és jelentőségét Csepeli György a következők szerint foglalta össze (Csepeli 2015: 173–174). (1) A megismerni kívánt aktoroktól származó viselkedés teljeskörűen vizsgálhatóvá válik (azaz  $n=all$ ), (2) az adatok keletkezése és elemzése közötti idő lerövidül, (3) a longitudinális megközelítések alkalmazásának lehetősége szélesedik (a társadalmi folyamatok időbeli alakulását is követni lehet), és (4) a kognitív szűrők, hipotézisek, kérdőívek hiánya miatt közvetlenül a szereplők viselkedését tudjuk megismerni (nem véleményeket). Az információ speciális keletkezési módjának köszönhetően az adatokat nem befolyásolja a kutatásra szánt pénzügyi források mérete, a megkérdezés módja, a kérdezőbiztos személye és a mintavétel esetleges korlátai.

Ha az új lehetőség adta módszer kétségeire vagyunk kíváncsiak, akkor Z. Karvalics László, Németh Renáta és Székely Iván írását feltétlen el kell olvasni. Z. Karvalics (2015: 200) a téma „logikai vakfoltjaként” három olyan pontot fogalmaz meg, melyekre a metaadatokkal dolgozó kutatóknak figyelemmel kell lenniük: a hermeneutizálás (az átcúszás veszélye a „mi lenne, ha” világába), a metonimizálás (egy kiválasztott rész alapján az egészre levont következtetés pontatlansága) és a dekontextualizálás (az adat-ökoszisztéma szerinti adaptáció fontossága). A szerző szerint a probléma az, hogy láthatjuk például az egyéni viselkedések különbözőségeit, de az okokra nem kapunk magyarázatot, mert az adatokból csak a következmények látszanak. Németh Renáta és Székely Iván tanulmányai már Dessewffy és Láng problémafelvető téziseire adott válaszok. Németh szerint az adatok csupán digitális lábnyomok, melyeknek „önmagukban nincsen jelentésük, a jelentést a kutató alkotja” (Németh 2015: 205). Írásában számos egyéb korlátra (a minta teljes populációval való azonosításának problémáira, a korrelációs minták és az oksági viszonyok tisztázására) hívja fel a figyelmet, melyeket folyamatosan szem előtt kell tartani a „nagy adatok” elemzése során. Székely Iván – tisztázandó kérdésekként – a Big Data gyenge pontjait összegezi. Ilyen például az adatelemzés kizárólagosságának módszertana, melyben Z. Karvaliccsal összhangban a miértekre adható válaszok hiánya a hangsúlyos. Az új típusú adatforrás-felhasználás határainak kijelölése során az adatvédelem szempontja kiemelt területként jelenik meg Székelynél. A tisztázandó

kérdések egyike és a tanulmány második része (adatmentes zónák) szinte teljes terjedelmében ebből a fókuszpontból vizsgálja a témakört.

A *Replika* 92–93. számának tematikus blokkjában egyedi tartalomként olvashatjuk Gayer Zoltán like-olási szokások alapján végzett klaszterelemzését, mely a gyakorlati felhasználhatóság esettanulmányaként értékelhető. Székely írásával annyiban érdemes szembeállítani, hogy Gayer épp az egyének vizsgálatában látja a Big Data által kivitelezhető kutatások jelentőségét, ami új lehetőség a kategóriák viselkedésének tanulmányozása mellett, és nem az alanyok személyes életterébe való kényszerű betekintés fóruma.

### Találkozás egy „igazi” Big Datával

PhD-kutatásom során folyamatosan kerestem adatokat önkormányzati képviselőkről a [valasztas.hu](http://valasztas.hu) honlapján, mely témánk szempontjából egy „open data” típusú adatbázisként definiálható. Bár az adatok viszonylag jól strukturálva, részletgazdagon, választási évek szerint elkülönítve, település- és sok esetben választókerületi szintig lebontva kereshetők, nagyobb (például több településre vagy évre kiterjedő reprezentatív) országos minta összegyűjtésére nehezen alkalmazható. Az általam vizsgálni kívánt téma, az önkormányzati képviselők arányainak alakulása a rendszerváltás utáni Magyarországon, hét választási évre vonatkozó adatainak összegyűjtése reménytelen küldetésnek tűnt a nyitott adatbázisból, ezért a Nemzeti Választási Iroda Választási Információs Szolgálatától adatigénylés formájában kikértem az önkormányzati választások eredményeit.<sup>1</sup> Kérésem alapján (ZIP fájlba) tömörítve 615 MB-nyi adat érkezett. Az adatok txt kiterjesztésűek, strukturáltak (17 800 file), szerkezetileg a választási jegyzőkönyvek felépítését követik (fej és tételsor szerinti bontásúak). A txt fájlok változó sor- és mezőhosszúságúak, win 1250 kódkészlettel készültek.

Dessewffy és Láng korábbi széles értelmezését használva a kutatás során elemzett választási adatok a hazai közigazgatásban keletkező, Big Data típusú „metaadat” körébe sorolhatók. Tudatos emberi tevékenység nyomán (Dessewffy és Láng 2015: 158) jöttek létre, de nem megfigyelésből és nem kutatói szándékból születtek. Keletkezésük alapja a demokratikus választások alkalmával, négyévente megrendezett önkormányzati megmérettetések „végtermékének” rögzítése, vagyis a választási eredmények digitalizált jegyzőkönyvei. Mivel ezek tartalma az önkormányzati képviselők teljes populációjára vonatkozik, relatív értelemben megvalósul a Big Data sajátossága. A „nagy adat” ugyanis nemcsak a terabájnyi kiterjedésű adatállományokat jelöli, hanem az elemzés tárgyát képező, a vizsgálni kívánt jelenséggel kapcsolatos összes információ felhasználását is (Mayer-Schönberger és Cukier 2012). Tehát a lényeg nemcsak az adatok mérete, hanem azok teljeskörűsége és a keletkező információ újrahasznosítása. A Big Data koncepciója szerint ugyanis az információ (esetünkben a választási eredmények) értéke jóval nagyobb, mint amire elsődlegesen használjuk (az egyes körzetek nyerteseinek megállapítása). Az adatok opciós értékkel rendelkeznek, mely az új célokra való felhasználás során keletkezik. Estünkben akkor, mikor az önkormányzati képviselők nemi megoszlását és mandátumszerzési sajátosságait vizsgáljuk a választási eredmé-

1 Ezúton mondok köszönetet témavezetőmnek, Dr. habil. Vámosi Tamásnak a folyamatos támogatásért, valamint a VISZ azon munkatársának, aki segítőkész információval ismeretlenül is hozzájárult a kutatás elindításához.

nyek alapján. A lehetőségek kiaknázásához a szükséges adatok megfelelő kiválasztására és kezelésére van szükség. Egyáltalán nem biztos, hogy minden adatsorra szükség van (Élő és Szármas 2015: 163–164).

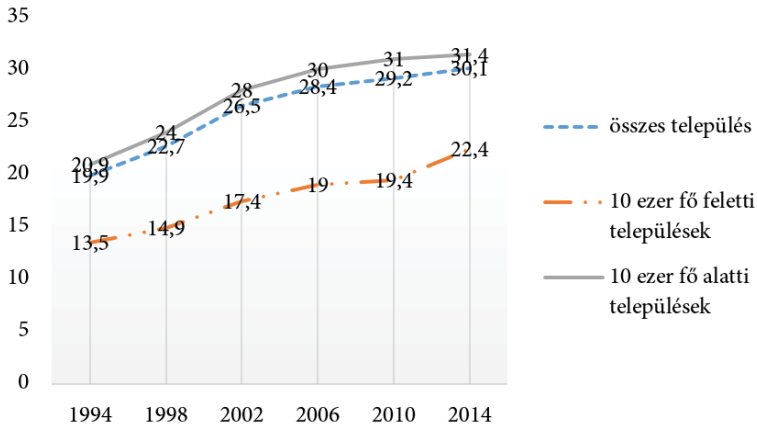
A NISZ által elküldött adatokat változatos tartalom és struktúra jellemzi, ezért az első találkozás velük maradandó kutatói élményt jelentett. A releváns információk kiválasztása a nagy mennyiségű adatokból, szűrésük és elemezhető formába rendezésük, összekapcsolásuk számos kihívást rejtett magában, sokszor választási évenként új megoldási utakat eredményezve. Ez azonban adattisztítással, kreativitással a kisebb problémák körébe tartozik. Az igazi korlátot nem ez jelentette, hanem a rögzített adatok tartalma és a 2-3 választásonként eltérő struktúrája. Mivel az adatgyűjtést, feldolgozást, rögzítést nem egy alaposan átgondolt kutatói szándék vezérelte, az adatok nagymértékben befolyásolják a kivitelezhető kutatás területeit és tematikáját (például a rendszerváltás utáni első önkormányzati választásra vonatkozóan a felhasznált adatbázis nem tartalmaz részletes információkat). Viszont kárpótlásként az eredmények a mandátumot nyert önkormányzati képviselők teljes populációjára vonatkoznak, vagyis a vizsgálandó célcsoport esetében ténylegesen igaz az  $n=all$ . Ez alapján kutatásom egy statisztikai adatokat ismertető alapkutatásnak tekinthető, mely – a megismerésen túl – alkalmas lehet arra, hogy kiinduló arányokat, ismereteket nyújtson további empirikus vizsgálatokhoz. Ebben a konkrét esetben az adatbázis lehetőségei és a kutatási téma alapján elvégezhető vizsgálat időhorizontja: az 1994–2014 közötti hat önkormányzati választás. Az elemzés célcsoportja: a mandátumot szerzett önkormányzati és megyei közgyűlési képviselők (polgármesterek és kisebbségi képviselők nélkül). A kutatás fókuszja: a helyi képviselők vizsgálata, nemi megoszlásra, mandátumismétlésekre és területi sajátosságokra való tekintettel. Az eredmények országos összesítésben és különböző területi szinteket összehasonlítva (10 ezer fő alatti településen, 10 ezer fő feletti városokban, egyéni választókerületi és kompenzációs listán mandátumhoz jutott képviselők, megyei közgyűlési tagok) modellezhetik a női részvételi arányok időbeli alakulását.

## Kutatói vélemények vs. választási adatbázis

### *Amit a választási adatok igazoltak*

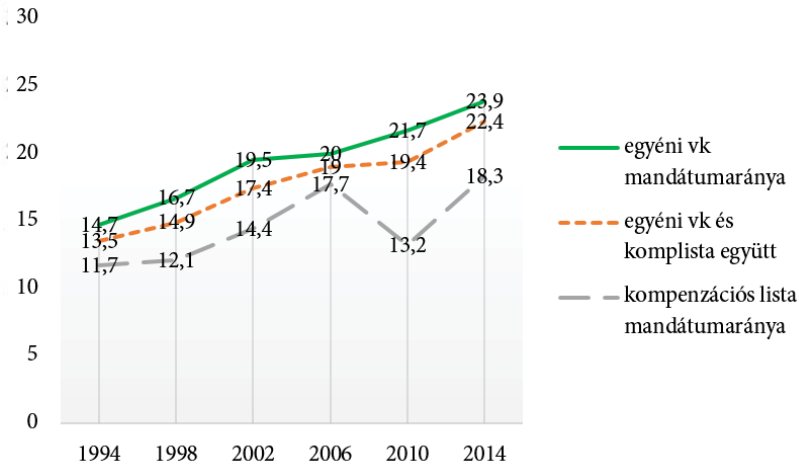
Az idézett *Replika*-tanulmányokban a Big Data definíciója során két fontos mozzanat került előtérbe: a megnövekedett adatmennyiség és a digitális rögzítési forma általánossá válása (Dessewffy és Láng 2015: 157). Ez utóbbi, vagyis az önkormányzati választási eredmények nem csak papíralapú dokumentálása és tárolása vitathatatlanul jellemző a használt adatbázisra. A megnövekedett információ a rögzített adatok részletgazdagságában ragadható meg. Nemcsak eredményeket, átlagokat, időbeli változást vizsgálhatunk a győztesek, mandátumot nyert képviselők kapcsán, hanem a különböző fájlteralmak összekapcsolásával összefüggések, egyedi sajátosságok vizsgálhatók. A női képviseleti arányok esetében (az első eredményeim szerint) például a mandátumszerzés módja (2. ábra) és a település mérete (1. ábra) egyértelműen befolyásoló tényező.

1. ábra. Női képviseleti arányok az önkormányzatokban



Forrás: választási adatok (NVI), saját szerkesztés

2. ábra. A képviselőnők aránya az önkormányzatokban a 10 ezer fő feletti településeken

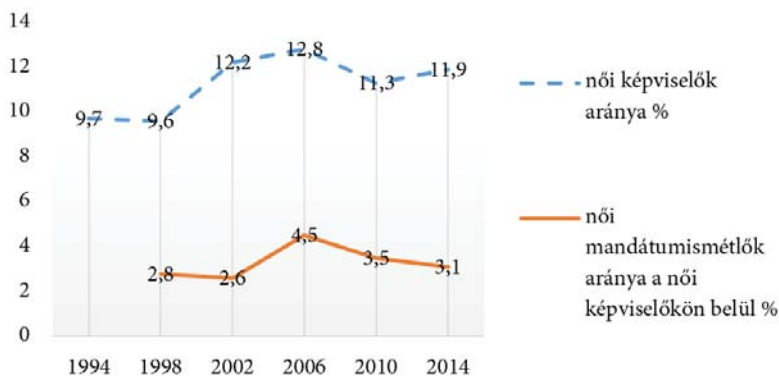


Forrás: választási adatok (NVI), saját szerkesztés

Dessewffy és Láng (2015: 159) Latourra hivatkozva (Latour 2010) a Big Data előnye-ként azt emelik ki, hogy az archivált digitális adatbankok alapján aggregát adatok (lásd Durkheim) és az individuális szint (lásd Tarde) egyaránt vizsgálható, ami érdekes és új kutatási eredményeket hozhat. Az általam használt választási adatokon ez szintén modellezhető. A jelöltek és képviselők név szerint szerepelnek a választási jegyzőkönyvekben, ezáltal például 1-1 képviselő mandátumismétlését (*inkumbens* képviselők vizsgálata, lásd Tóth és Ilonszki 2015: 35) is követni tudjuk, ami például a helyi politikai elit összetételének,

a professzionális politikusi csoportok kialakulásának vizsgálatához jól használható. A kutatás során tehát nem csak országosan vagy megyei szinten összesített arányokat tudok megmutatni, hanem az egy-egy településre vagy önkormányzati képviselőre vonatkozó egyéni információ is nyomon követhető.

3. ábra. A megyei közgyűlések női mandátumai

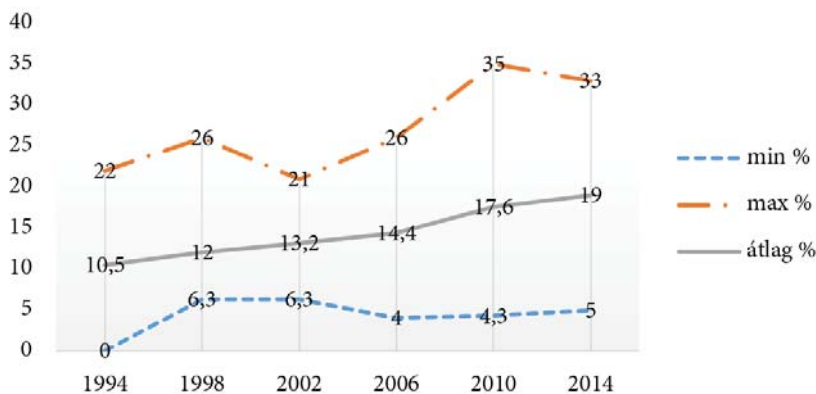


Forrás: választási adatok (NVI), saját szerkesztés

A 3. ábra a megyei közgyűlések női képviselőinek részvételi arányait és ezen belül azon nők százalékát mutatja, akik legalább két egymást követő alkalommal képviselői mandátumot szereztek. Ezen túlmutatóan az is látszik az adatbázisban, hogy 2002-ben 8, 2006-ban 5 olyan képviselőnő vett részt országosan közgyűlési munkában, aki harmadszor (2002) és negyedszer (2006) is részt vállalt a megye életének irányításában (ez már a személyes szint vizsgálata). Egy interjú kutatás (pl. a női politikai elit vizsgálata kapcsán) során érdemes lenne megkeresni őket.

A Big Data használatának egyik legvonzóbb jellemző sajátossága az  $n=all$ , vagyis a minta méretének egybeesése a populáció nagyságával (Mayer-Schönberger és Cukier 2013). A teljeskörűség összetett kérdés, több irányba mutat, és a tematikus kötet szerzői is eltérő következtetésekre jutottak ezzel összefüggésben. Ezeket követve az első neuralgikus pont a „Tényleg  $n=all$ ?” kérdése. Kutatási eredményeim alapján a választási adatbázis esetében igazolódik a tézis. A választási eredmények alapján kiosztott önkormányzati mandátumok a tényleges képviselői populáció egészét mutatják – szemben a Németh Renáta (2015: 204) által megfogalmazott példával (a Facebook- és Twitter-bejegyzések elemzése), ahol helytálló az az észrevétel, hogy a vizsgálatba a populációnak csak azon köre kerül be, amelyik az adott időintervallumban felhasználója a felületeknek. A „teljeskörűség” kapcsán fogalmazódik meg az is, hogy a „nagy adatok” által a valódi struktúrák és ezek egymáshoz való viszonya is vizsgálható. Csoportok megismerésekor (Csepeli 2015: 174) a csoportképzést nem előzetes kutatói megfontolások alakították, hanem az aktorok tényleges magatartása. A szabályszerűségek megismerése után pedig a kivételek és szokatlan együtt járárok is vizsgálhatók a nagy adattömegben (ahogy ezt a következő ábra is mutatja).

4. ábra. A megyei jogú városok nőimandátum-aránya



Forrás: választási adatok (NVI), saját szerkesztés

A 4. ábrán azt láthatjuk, hogy a megyei jogú városokban megfigyelhető, összességében lassan emelkedő tendenciájú női mandátumarány átlagához képest az adott éveket tekintve mekkora eltérések, szélső értékek tapasztalhatók. Az utolsó két választás alkalmával meglehetősen nagyra nőtt a különbség a minimum és maximum százalékos értékei között (30–31%). Egy okokat feltáró empirikus kutatás során érdemes lehet megvizsgálni a két kiugró eredmény mögött álló folyamatokat.

Végezetül a hagyományos kérdőívekkel, módszerekkel szemben a Big Data-adatbázisok előnyeként fogalmazódik meg (Dessewffy és Láng; Csepeli) a valódi viselkedési minták („digitális lábnyomok”) mérésének lehetősége a vélemények helyett. Adatbázisom esetében a választói aktivitás területi szintjeinek különbségei, a pártpreferenciák vagy a női politikusokkal kapcsolatos attitűdök (például a szavazási hajlandóság) változása is vizsgálható. A nők önkormányzati döntéshozatalba, politikai pozíciókba kerülési esélyeit kutatva kifejezetten előnyösnek tartom, hogy a válaszadókat nem befolyásolta egy esetleges (például a női interjúkészítő személyének címzett) megfelelési vágy.

#### *Ami nem vagy korlátozottan jellemzi a választási adatbázist*

Dessewffy és Láng tanulmányukban hivatkoznak Anderson 2007-ben megfogalmazott koncepciójára (*Az elmélet vége*), mely szerint „az elmélet által vezérelt kutatásokból az adatok által vezérelt vizsgálatok felé” (Dessewffy és Láng 2005: 161) tart a tudomány. A vizsgált adatbázis tekintetében nyomokban tapasztalható, hogy az adatok irányítanak (pl. 1990-es választási részletes eredmények hiánya esetében), de összességében egyetértek Dessewffy és Láng megállapításával, akik abszurdnak nevezik az „elmélet végét”. Vizsgálatom esetében az egyéni kutatói érdeklődés inspirálta a téma vizsgálatát (az elméleti koncepció megalkotását), majd a feldolgozás reprezentatív igényű kivitelezési hozzádoka volt az adatbázis felkutatása és használata. Nem pedig fordítva. Hasonló következtetésre jutottam James C. Scott – szín-

tén a szerzőpáros által idézett – ama véleményének „szembesítése” során is (Dessewffy és Láng 2015: 161), mely szerint a nemzetállamok bürokratikus kontrollfunkciója és a folyamatok adatosítása (*datafication*) húzódik a standard adatkezelési szokások kialakítása mögött. A választási eredmények, adatok nyilvánossága, folyamatos elérhetősége épp az állampolgári kontrollt és az átláthatóságot biztosítja és nem az orwelli nagy testvért erősíti.

Az időbeliség (Dessewffy és Láng 2015: 166) vagy azonnalóság (Csepeli 2015: 173) érvényessége esetünkben nagyon speciális. A választási eredmények első és döntő része még aznap megszületik, de a közigazgatás egy olyan szintjén, mely nem kutatható azonnal. A Big Data előnyeként számon tartott longitudinalitás (Csepeli 2015: 174) sem feltétlen jellemzi a felhasznált adatbázist. Az önkormányzati szereplők vizsgálatának megközelítési alapját a betöltött funkciójuk képezi, vagyis a mandátumot nyert képviselők mint személyek nem képeznek panelmintát, sok a cserélődés közöttük, mint ahogy a választópolgárok csoportjaiban is nagy a fluktuáció. Tehát a Big Data e két utóbbi sajátossága az általam használt esetben nem igazolható.

Végezetül talán az egyik legizgalmasabb szempont maradt hátra: láthatjuk-e a jövőt az általunk használt választási adatok elemzésével? Megjósolható-e bármi a következő képviselő-testületek összetételére, meg tudjuk-e mondani, hogy ki nyer a következő választáson? Ami tény: önmagában a nagy minta (elemszám) nem garantálja a közvélemény-kutatások előrejelzései alapján megfogalmazott eredménybecslések pontosságát (Rudas 2006: 2). Márpedig a választók szempontjából sosem lesz  $n=all$ , így erre a perspektívára a felhasznált adatbázis nem szolgáltat jövőbeli információt. A választói magatartás ugyanis elvileg nem előre jelezhető. (A gyakorlatban természetesen vannak választási előrejelzések, becslések, de hogy milyen adatokból – közvélemény-kutatás alapján, korábbi eredmények földrajzi eloszlásának vizsgálatából – és milyen módszerrel kerülnek kiszámításra, az már egy más témakörbe tartozó módszertani probléma (lásd Karácsony és Lakatos 2006: 186–190; Helmich és Szántó 2004). A kutatás női arányokat elemző szempontja alapján viszont a trend megállapítható a felhasznált adatbázisból (mely tulajdonképpen nem más, mint az adatelemzések összefoglaló eredménye). Azaz főszabályként „előre látható”, hogy egy női képviselőjelöltből a 10 ezer fő alatti településeken és a 10 ezer főt meghaladó lakosságszámmal rendelkező városok egyéni választókerületeiben lesz a legnagyobb eséllyel helyi politikus. Ugyanerre a megyei önkormányzatok közgyűlésében sokkal kisebb az esélye. Ebből két fontos dolog következik. Az egyik: a magasabb arányú női képviselő döntően ma már nem a választópolgárokon múlik, a nemzeti tendenciáknak megfelelően (Tóth és Ilonszki 2015: 30) Magyarországon is változik a szavazási hajlandóság a női jelöltekre vontakozóan. A másik: a helyi, önkormányzati politikában is (elsősorban a 10 ezer fő feletti lakosságszámmal rendelkező városok esetében) megfigyelhető az a pártselekción mechanizmus (a szervezetek kapuőrszerepe, lásd Várnagy és Ilonszki 2012), melyet az országgyűlés esetében már korábban megismerhettünk.

## Összegzés

„Az internet behatol a társadalom alrendszerébe. Az internet közegébe költözik az üzleti élet, a politika, a kormányzás, az oktatás, a gyógyítás. Az életvilág, mely korábban a magánélet kitüntetetten intim, zárt köre volt, az interneten kinyílik a végtelenbe a megerősítésekre éhes én új terepeként” – írta Csepeli György (2015: 172). Az önkormányzati döntéshozatal



szereplőivel is ez történt, amikor a választási jegyzőkönyvek digitalizálása megkezdődött és egy internetes felületnek köszönhetően bárki számára elérhetővé, megismerhetővé vált. Ha a Big Data kapcsán a pró és a kontra oldal közül kell választani, akkor inkább a „nagy adatok” elemzését üdvözlő kutatókkal értek egyet, de azt ismételten hangsúlyozom, hogy tapasztalataim egyetlen adatbázis használatából származnak. Ha módszertani szempontból közelítjük meg a kérdést, akkor a minta és a populáció egybeesése rendkívül izgalmas és új kutatói perspektívát, eredményeket hoz magával. Az a vélemény viszont tovább is erős maradt bennem, hogy az adatok „vallatása” során a megismerés egy nagyon fontos szintjéig (az alapok feltárásáig, összefüggések megállapításáig) el tudunk jutni, mely után empirikus kutatások válnak szükségessé, lehetőséget biztosítva a tendencia okainak megismerésére. E kettő együttese vihet minket közelebb a társadalmi folyamatok mélyebb értelmezéséhez.

### Hivatkozott irodalom

- Anderson, Chris (2007): The End of Theory. The Data Deluge Makes the Scientific Method Obsolete. *Wired* 16(7). Interneten: <https://www.wired.com/2008/06/pb-theory/>.
- Csepeli György (2015): A szociológia és a Big Data. *Replika* (92–93): 171–177.
- Desseffy Tibor és Láng László (2015): Big Data és a társadalomtudományok véletlen találkozása a műtőasztalon. *Replika* (92–93): 157–170.
- Élő Gábor és Szármas Péter (2015): A big data technológia gazdasági és társadalmi jelentősége. In *Metszéspontok. Társadalomtudomány és infokommunikáció az ezredforduló után*. Z. Karvalics László (szerk.). Budapest: Gondolat, 159–216.
- Gayer Zoltán (2015): Ízléselemzés és Big Data. *Replika* (92–93): 177–189.
- Helmich Dezső és Szántó Zsolt (szerk.) (2004): *Metodológia, társadalom, gazdaság. In memoriam Bertalan László*. Budapest: Educatio. Interneten: <http://www.tankonyvtar.hu/en/tartalom/tkt/metodologia-tarsadalom/ch28s02.html> (leölve: 2017. május 31.).
- Karácsony Gergely és Lakatos Zsuzsa (2006): Választási előrejelzések Magyarországon, 1990–2006. Sikerek, tévedések és tanulságok. In *Mérésről mérésre. A választáskutatás módszertani kérdései, 2006*. Angelusz Róbert és Tardos Róbert (szerk.). Budapest: DKMKA, 185–217. Interneten: [http://www.median.hu/kepek/upload/2010-01/karacsony\\_lakatos\\_2006.pdf](http://www.median.hu/kepek/upload/2010-01/karacsony_lakatos_2006.pdf) (letölve: 2017. június 3.).
- Latour, Bruno (2010): Tarde's Idea of Quantification. In *The Social after Gabriel Tarde. Debates and Assessments*. Matei Candea (szerk.). New York: Routledge, 143–163.
- Mayer-Schönberger, Viktor és Kenneth Cukier (2013): *Big Data. A Revolution That Will Transform How We Live, Work and Think*. London: John Murray.
- Németh Renáta (2015): A számok tényleg magukért beszélnek? *Replika* (92–93): 203–209.
- Rudas Tamás (2006): *Közvélemény-kutatás*. Budapest: Corvina. Interneten: [http://www.tankonyvtar.hu/en/tartalom/tamop425/2011\\_0001\\_656\\_kozvelemenyn\\_kutatas/2011\\_0001\\_656\\_kozvelemenyn\\_kutatas.pdf](http://www.tankonyvtar.hu/en/tartalom/tamop425/2011_0001_656_kozvelemenyn_kutatas/2011_0001_656_kozvelemenyn_kutatas.pdf) (leölve: 2017. május 31.).
- Székely Iván (2015): Az adatmentes zónák szükségessége és esélye. *Replika* (92–93): 209–227.
- Tóth Adrienn és Ilonszki Gabriella (2015): Pártok vagy választók? A női jelöltek esélye az egyéni választókerületekben, 1998–2010. *Politikatudományi Szemle* 24(3): 27–50.
- Várnagy Réka és Ilonszki Gabriella (2012): Üvegplafonok. Pártok lent és fent. *Politikatudományi Szemle* 21(4): 7–25.
- Z. Karvalics, László (2015): A Nagy-Adat jelenség társadalomtudományi lehorgonyzásához. *Replika* (92–93): 189–203. [www.valasztas.hu](http://www.valasztas.hu)